BindingDB

A Publicly Accessible Database for Drug Discovery

Michael K. Gilson, M.D., Ph.D.
Skaggs School of Pharmacy and Pharmaceutical Sciences
U. C. San Diego

NIST-Rutgers Workshop 1997 & NSF Proposal 1998

"The First Public, Web-Accessible Database
with Information on Molecular Recognition"

MKG, Bourne, Westbrook

K, IC50

$$A + B \rightarrow A \bullet B$$

Experimental conditions, other associated data

Protein-ligand, protein-protein, nucleic acids, host-guest

Binding Databases: Important Components of the Bioinformatics Infrastructure

- Accommodate data sets too large to be published
- Accommodate raw data, making reanalysis possible
- Facilitate lead compound discovery
- Validate and improve computer-aided drug design software
- Flag risk of side-effects, based on chemical similarity
- Elucidate mechanisms of bioactive compounds
  Identify biomolecules bound by chemical similars

“The database is now up and running…”

2000-2001

“The next key task is to gather data. This proves to be somewhat challenging…”

JOURNAL OF MOLECULAR RECOGNITION
J. Mol. Recognit. 2002; **15**: 1
DOI:10.1002/jmr.560

Editorial

The bioinformatics of molecular recognition

The establishment of the major gene and protein databases, and the imperative to interpret and use the data they contain, have created the field of bioinformatics. However, while the major biomolecular databases focus almost exclusively on single molecules, it is fundamentally the *interactions* among biomolecules that endow them with function—a fact that needs no emphasis in the pages of *Journal of Molecular Recognition*.

It is thus significant that, in the last few years, databases have finally been established to archive and present data on molecular binding interactions. Two of these databases—the Database of Interacting Proteins (DIP; http://dip.doe-mbi.ucla.edu/) and the Biomolecular Interaction Network Database (BIND; http://www.bind.ca)—focus on listing binding interactions among naturally occurring molecules, as determined by 2-hybrid studies, for example. However, readers may be more interested in the *quantitative* binding data in the Thermodynamic Database for Protein-Nucleic Acid Interactions (ProNIT; http://www.rtc.riken.go.jp/jou-hou/pronit/pronit.html) and this author's Binding Database (BindingDB; http://www.bindingdb.org). Both data collections are young but growing. ProNIT, which is limited to protein–nucleic acid binding, currently lists *ca*2000 binding interactions involving 92 different proteins. BindingDB, which collects data for all kinds of molecules and complexes, including synthetic compounds and mutant biopolymers, currently lists *ca*250 binding interactions involving about 100 different molecules.

What is a molecular recognition database good for? A number of uses are readily envisioned. For example, after identifying a protein of interest in GenBank or SwissProt, it is natural to ask what other molecules it binds. In cases where one wishes to understand a biological system better, one will be interested in interactions with naturally occurring proteins or nucleic acids. For drug design, small organic ('drug-like') compounds may be more important. For example, an extensive binding database could facilitate the discovery of lead compounds by allowing one to quickly identify small molecules that bind homologs of a targeted protein. Conversely, if the database shows that a drug candidate binds many different proteins, this might indicate a high risk of side-effects. And for theoreticians, ready access to a range of binding data would facilitate the development of quantitative models of binding for use in molecular modeling and drug design.

In order to support all of these activities, BindingDB collects binding data on both naturally occurring and synthetic compounds. Affinities of course depend upon experimental conditions, so BindingDB stores not only the measured affinities themselves, but also pH, temperature and buffer composition, along with materials and methods that are important for reproducing and assessing the data. Currently, BindingDB contains measurements by isothermal titration calorimetry only, but measurements by enzyme inhibition assays should be available soon—possibly by the time this piece is printed—and other techniques will be accommodated in the future. Measurement techniques are added one by one because the database must be expanded to accommodate each technique's special features. (For example, enzyme-inhibition methods require that the substrate be specified.)

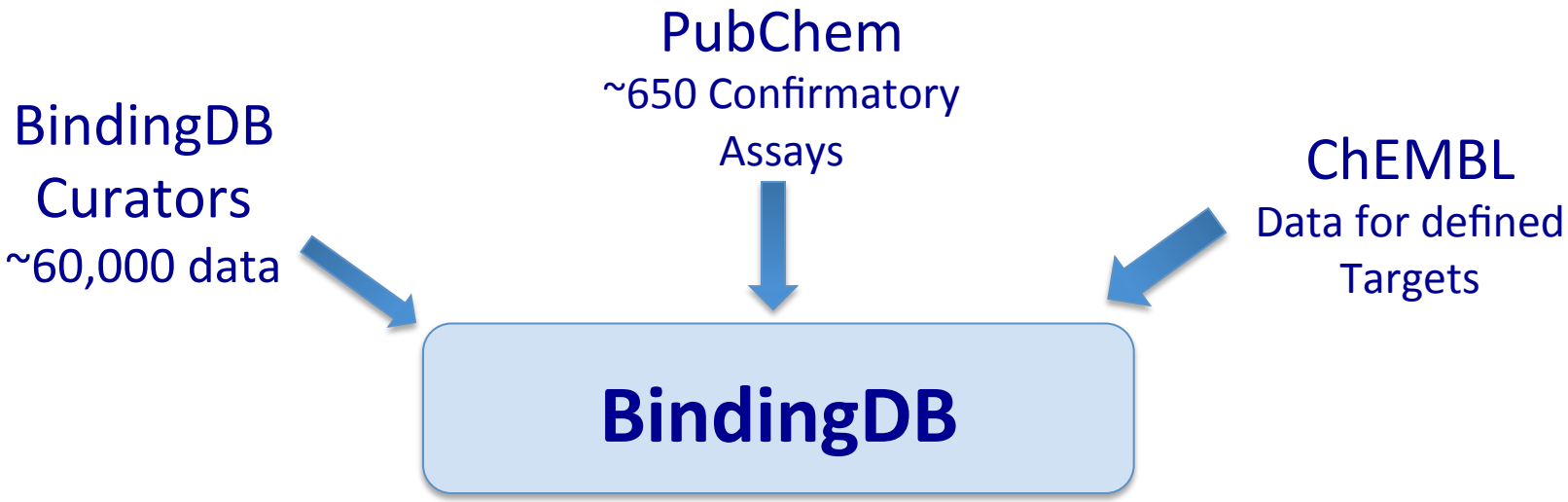
The database is now up and running, providing a range of search methods, including searches by keyword, author, molecule name, protein sequence homology and chemical similarity and substructure. The next key task is to gather data. This proves to be somewhat challenging—certainly more so than collecting DNA sequences, which are poured out in electronic format by automated instruments. In contrast, most molecular recognition data are locked away in journal articles where they can be read by humans but not readily parsed into database fields by a computer. The problem is being addressed in two ways. First, authors are invited to enter their binding data via the on-line forms provided at the database's web-site. The only restriction is that the methods used to generate the data be published in a refereed journal. Note that, although the forms request a detailed description of the measurements, the author who is pressed for time need enter only the relatively few mandatory items, which are marked as such on the data entry forms. Also, data can be corrected at any time by the author, even after they are in the database and available on the web. Depositors should feel free to ask BindingDB staff for help with the deposition process. Second, software tools are being developed that will speed the recovery of binding data from articles that are already published, but it will take some time before these methods are operational.

Future plans for BindingDB include the creation of more sophisticated search and analysis tools at the web-site and integration with other biomolecular databases. For example, it should be possible to link directly from GenBank sequences to BindingDB in order to discover compounds that bind a protein of interest; and correlations between structure and energetics should be supported via links with the Protein Data Bank (http://www.pdb.org) and with the protein–ligand database Relibase (http://relibase.ebi.ac.uk). Given the progress already made in establishing BindingDB and other molecular interaction databases and the importance of molecular recognition in biology, pharmaceutics and chemistry, binding data appear on track to become an important component of the bioinformatics infrastructure.

Michael K. Gilson
Center for Advanced Research in Biotechnology
University of Maryland
Biotechnology Institute
Rockville MD 20850
USA

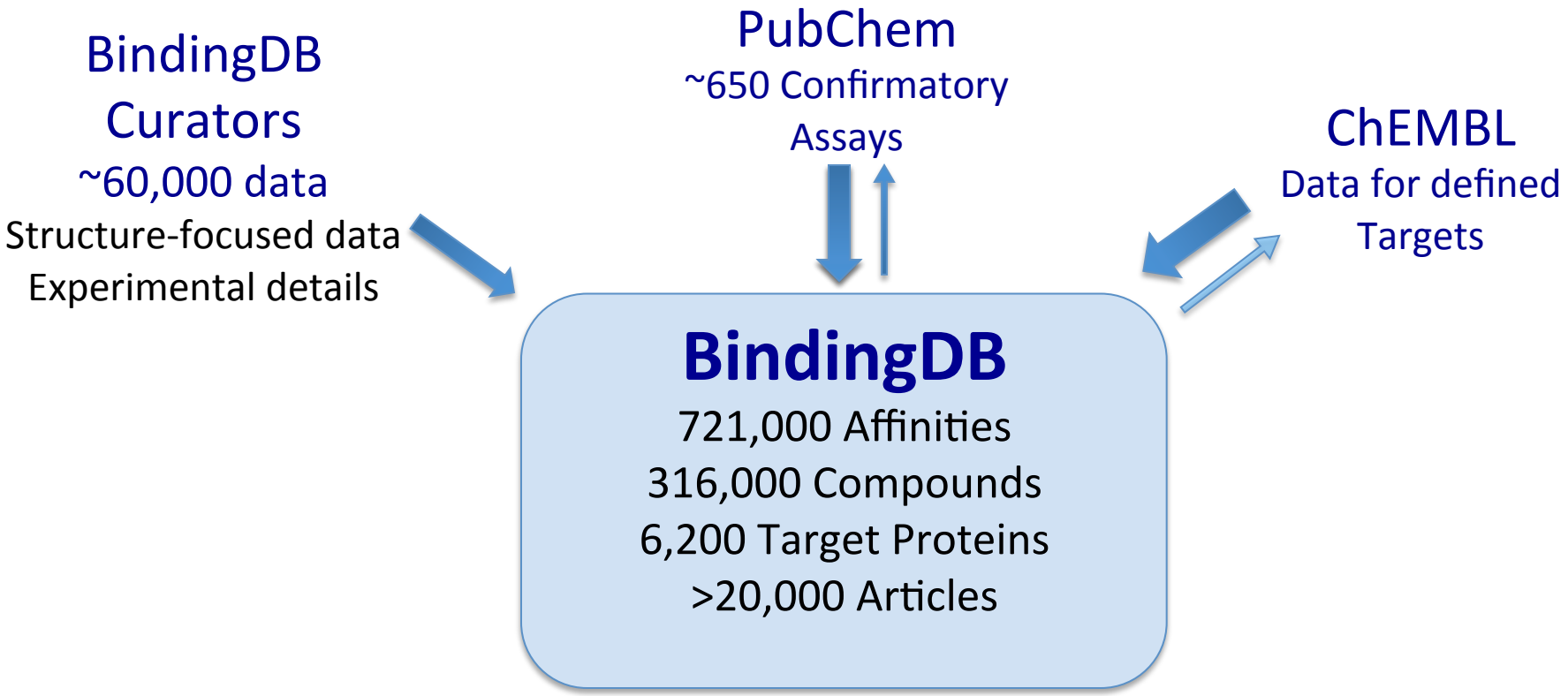
Data Collection

Measured Protein-Small Molecule Affinities



Data Collection

Measured Protein-Small Molecule Affinities



Serving a Varied Scientific Community

Didactic Material, Tutorials and Documentation

Info Page

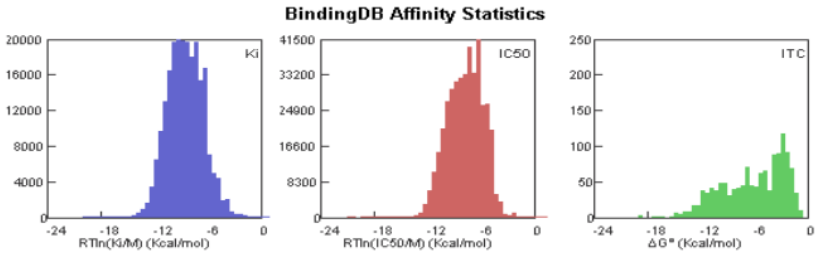
Info

BindingDB is a public, web-accessible database of measured binding affinities, focusing chiefly on the interactions of proteins considered to be candidate drug-targets with ligands that are small, drug-like molecules. BindingDB supports medicinal chemistry and drug discovery via literature awareness and development of structure-activity relations (SAR and QSAR); validation of computational chemistry and molecular modeling approaches such as docking, scoring and free energy methods; chemical biology and chemical genomics; and basic studies of the physical chemistry of molecular recognition. BindingDB also includes a small collection of host-guest binding data of interest to chemists studying supramolecular systems.

The data collection derives from a variety of measurement techniques, including enzyme inhibition and kinetics, isothermal titration calorimetry, NMR, and radioligand and competition assays. BindingDB includes data extracted from the literature by the BindingDB project, selected PubChem confirmatory BioAssays, and ChEMBL entries for which a well defined protein target ("TARGET_TYPE='PROTEIN'") is provided. Data extracted by BindingDB typically includes more details regarding experimental conditions, etc. BindingDB currently contains about 620,000 binding data for 5,500 proteins and over 270,000 drug-like molecules. Data from ChEMBL is provided under a Creative Commons Attribution-Share Alike 3.0 Unported License.

- Background on Proteins, Small Molecules, and Binding
- How to Use BindingDB
- Glossary
- Search Templates in BindingDB
- Documentation of BindingDB's SDfile format
- A caveat regarding protein Target sequences
- Important information about a BindingDB SDfile
- BindingDB Page on Wikipedia

BindingDB Affinity Statistics

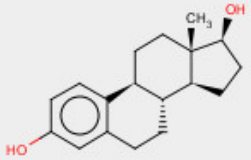


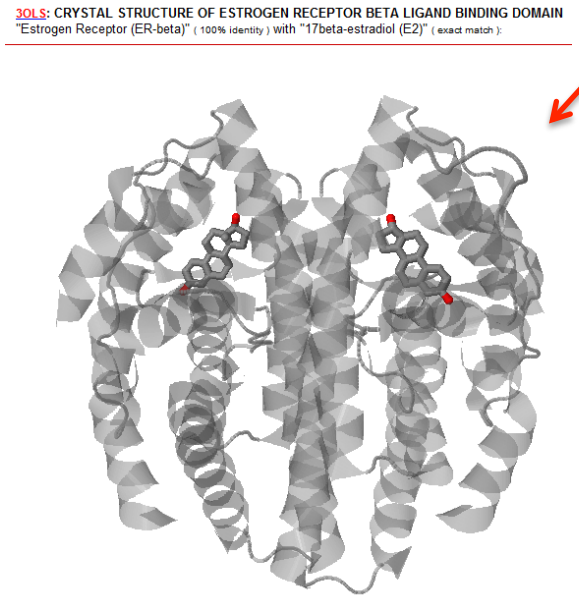
References

1. Liu,T., Lin,Y., Wen,X., Jorrisen, R.N. and Gilson,M.K. *BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities* Nucleic Acids Research *35:D198-D201* (2007). [pdf]
2. Chen,X., Lin,Y. and Gilson,M.K. *The Binding Database: Overview and User's Guide* Biopolymers Nucleic Acid Sci. *61:127-141* (2002).
3. Chen,X., Lin,Y., Liu,M. and Gilson,M.K. *The Binding Database: Data Management and Interface Design* Bioinformatics *18:130-139*(2002).
4. Chen,X., Liu,M., and Gilson,M.K. *Binding DB: A web-accessible molecular recognition database* J. Combi. Chem. High-Throughput Screen *4:719-725* (2001).

Linking Affinity to Structure

With Drs. Phil Bourne, Peter Rose et al., UCSD and Rutgers

Target (Institution)	Ligand	Target Links	Ligand Links	Trg + Lig Links	Ki nM	ΔG° kJ/mole	IC50 nM	Kd nM	EC50/IC50 nM	k_{off} s^{-1}	k_{on} $M^{-1}s^{-1}$	pH	Temp °C
Estrogen Receptor (ER-beta) (240/240 = 100%)† (Homo sapiens) GlaxoSmithKline Research Curated by ChEMBL	17beta-estradiol (E2)  ((1S,10R,11S,14S,15S)-15-methyltetracyclo[8.7.0.0^{...})	PDB MMDB NCI pathway Reactome pathway UniProtKB/SwissProt UniProtKB/TrEMBL B.MOAD DrugBank GoogleScholar	B.MOAD CHEBI DrugBank KEGG MMDB PC cid PC sid PDB	DrugBank MMDB PDB PubMed	2.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
					Bioorg Med Chem Lett. 18: 5075-7 (2008) 								



Linking Structure to Affinity

PDB

Summary | Sequence | Annotations | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | Methods

Crystal structure of estrogen receptor beta ligand binding domain

DOI:10.2210/pdb3ols/pdb

Primary Citation

Synthesis and crystal structure of a phosphorylated estrogen receptor ligand binding domain.

Mocklinghoff, S., Rose, R., Carraz, M., Visser, A., Ottmann, C., Brunsveld, L.

Journal: (2010) Chembiochem **11:** 2251-2254

PubMed: 20922740
DOI: 10.1002/cbic.201000532
Search Related Articles in PubMed

PubMed Abstract:

No abstract available... [Read More & Search PubMed Abstracts]

Molecular Description Show

Source Show

Related PDB Entries Show

Ligand Chemical Component Hide

Identifier	Formula	Name	Interactions
EST Search Download	$C_{18}H_{24}O_2$	ESTRADIOL	Ligand Explorer

External Ligand Annotations Hide

Identifier	Binding Affinity data from BindingDB	Binding Affinity data from BindingMOAD
EST Search Download	EC50: 0.039 - 30 nM IC50: 1 - 1230 nM Kd: 0.5 nM Ki: 2 - 100 nM	N/A

External Domain Annotations Hide

- PFAM Classification: 4 Domains - data from PFAM

Sourcing BindingDB Compounds

with Dr. John Irwin, UCSF

myBDB logout

- **Search and Browse**
  - **Target**
    - Sequence
    - Name &
      - Ki IC50 Kd EC50
      - ΔG° ΔH° -TΔS°
      - pH (Enzymatic Assay)
      - pH (ITC)
      - Substrate or Competitor
      - Compound Mol. Wt.
      - Chemical Structure
    - Source Organism
    - Number of Compounds
    - Monomer List in csv
    - Het List in SDF
  - **Compound**
    - FDA Drugs
    - Chemical Structure
    - Batch Structures
    - Name
    - SMILES
    - Number of Data / Targets
  - **Data Source**
    - Author
    - Journal/Citation
    - Institution
    - PubMed
    - PubChem BioAssay
  - **Special Data Sets**
    - Host Guest Systems
    - Prot-Lig Validation Sets
  - **Other Databases**
    - PDB 85% 100% Seq ID
    - UniProtKB/Swiss-Prot
    - UniProtKB/TrEMBL

Home | Info | Download | About us | Email us

Compile Data Set for Download or QSAR

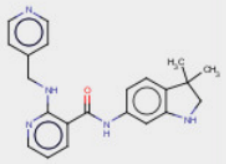
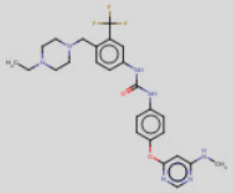
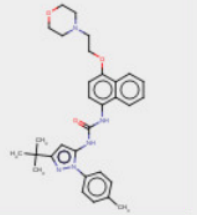
(2D ◉ Computed 3D by Vconf -m prep ○)

Add this page | Add all pages | Clear Selection | Make Data Set

E-MAIL

Found 106 hits Enz. Inhib. hit(s) with Target = 'ABL1'

Zinc 0: unavailable per Zinc DB. Zinc 1: purchasable, 2 weeks to supply. Zinc 2: made on demand. Zinc 4: potentially available

Target (Institution)	Ligand	Target Links	Ligand Links	Trg + Lig Links	Ki nM	ΔG° kJ/mole	IC50 nM	Kd nM	EC50/IC50 nM	K_{off} s^{-1}	K_{on} $M^{-1}s^{-1}$	pH	Temp °C
ABL1 (Homo sapiens)	cid_11667893  (AMG-706)	PDB MMDB NCI pathway Reactome pathway KEGG UniProtKB/SwissProt UniProtKB/TrEMBL B.MOAD DrugBank GoogleScholar	MMDB PC cid PC sid PDB ZINC 1 Similars	PCBioAssay	n/a	n/a	n/a	3900000	n/a	n/a	n/a	n/a	n/a
Ambit Biosciences					PubChem Bioassay (2008)								
ABL1 (Homo sapiens)	cid_11409972  (AST-487)	PDB MMDB NCI pathway Reactome pathway KEGG UniProtKB/SwissProt UniProtKB/TrEMBL B.MOAD DrugBank GoogleScholar	PC cid PC sid ZINC 0	PCBioAssay	n/a	n/a	n/a	5600	n/a	n/a	n/a	n/a	n/a
Ambit Biosciences					PubChem Bioassay (2008)								
ABL1 (Homo sapiens)	cid_156422  (BIRB-796)	PDB MMDB NCI pathway Reactome pathway KEGG UniProtKB/SwissProt UniProtKB/TrEMBL B.MOAD DrugBank GoogleScholar	DrugBank MMDB PC cid PC sid PDB ZINC 1 Similars	PCBioAssay	n/a	n/a	n/a	3400000	n/a	n/a	n/a	n/a	n/a
Ambit Biosciences					PubChem Bioassay (2008)								

BindingDB Compound Found in ZINC

UCSF University of California, San Francisco | About UCSF | Search UCSF | UCSF Medical Center — Shoichet Laboratory — docking.org

Search Results

A free database for virtual screening

ZINC is not commercial

Home | Subsets | HELP! | Mailing Lists

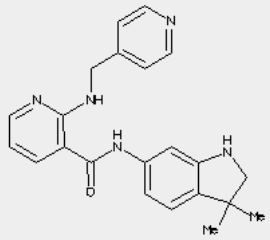
Initial search took 1 second(s).

Beginning : #1 Hits: 1 Display: per page= 100

Next Page | Download Table | Purchasing Info

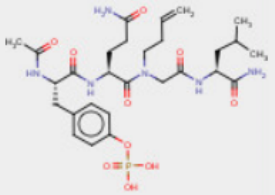
Download: SMILES | MOL2 | SDF | Flexibase | Show Detail?

Done

ZINC ID	Supplier information; Representations **Properties:** xLogP, ap & p desolvation, HBD,HBA,Charge,Mwt,NRB Annotations; Similarity	Structure Click for quick 3D display
18710082 #1 FIND SIMILAR Go SEA!	Selleck Chemicals: S1032 Toronto Research Chemicals: M732500 BindingDB.org: 24773, 31101 ChEBI: CHEBI:51098 ChEMBL-08: CHEMBL572881...17 total ref: mol2, SDF, SMILES, flexibase 2.65,7.36,-12.53,3,6,0, 373.46, 5 Tyrosine-protein kinase receptor FLT3: 1100 (ChEMBL09) Vascular endothelial growth factor receptor 3: 9.7 (ChEMBL09) Tyrosine-protein kinase FRK: 99 (ChEMBL09) Serine/threonine-protein kinase 2: 1300 (ChEMBL09) Vascular endothelial growth factor receptor 2: 26 (ChEMBL09) Mixed lineage kinase 7: 8 (ChEMBL09) Tyrosine-protein kinase ABL: 3900 (ChEMBL09) Stem cell growth factor receptor: 3.7 (ChEMBL09) Serine/threonine-protein kinase TNNI3K: 220 (ChEMBL09) Tyrosine-protein kinase FYN: 2800 (ChEMBL09) Eukaryotic translation initiation factor 2-alpha kinase 4: 4400 (ChEMBL09) Tyrosine-protein kinase LCK: 360 (ChEMBL09) ... 31 in total. Similar to: 34354825, 38239307, SEA predictions: Citron Rho-interacting kinase pPvalue:22 maxTC:100 (annotated) Epidermal growth factor receptor erbB1 pPvalue:0 maxTC:100 (annotated) Serine/threonine-protein kinase RAF pPvalue:2 maxTC:100 (annotated) Tyrosine-protein kinase LCK pPvalue:0 maxTC:100 (annotated) Macrophage colony stimulating factor receptor pPvalue:8 maxTC:100 (annotated) ... 26 in total.	

Toward a Biologist's View of Binding Data

Bridging to Systems Biology

Target (Institution)	Ligand	Target Links	Ligand Links	Trg + Lig Links	Ki nM	ΔG° kJ/mole	IC50 nM	Kd nM	EC50/IC50 nM	k_{off} s^{-1}	k_{on} $M^{-1}s^{-1}$	pH	Temp °C
SHC-transforming protein 1 (Homo sapiens)	CHEBI:618062  (4-((S)-2-acetamido-3-((S)-5-amino-1-((2-((S)-1-ami...)	PDB MMDB NCI pathway Reactome pathway KEGG UniProtKB/SwissProt UniProtKB/TrEMBL GoogleScholar	CHEBI KEGG PC cid PC sid	PubMed	n/a	n/a	6000.0	n/a	n/a	n/a	n/a	n/a	n/a
National Cancer Institute Curated by ChEMBL					*J Med Chem. 52: 1612-8* (2009)								

Pathway Interaction Database

National Cancer Institute/Nature

NATIONAL CANCER INSTITUTE
nature PathwayInteractionDatabase

Home > Browse pathways > NCI-Nature curated pathway

IL2 signaling events mediated by STAT5

Revision date: 12-Apr-2010
Curated by: Kira Anthony
Reviewed by: Angelita Rebollo, Hodaka Fujii, Massimo Gadina, Sarah L. Gaffen
Pathway ID: il2_stat5pathway
Pathway category: Interleukin mediated signaling pathways
Parent pathway: IL2-mediated signaling events

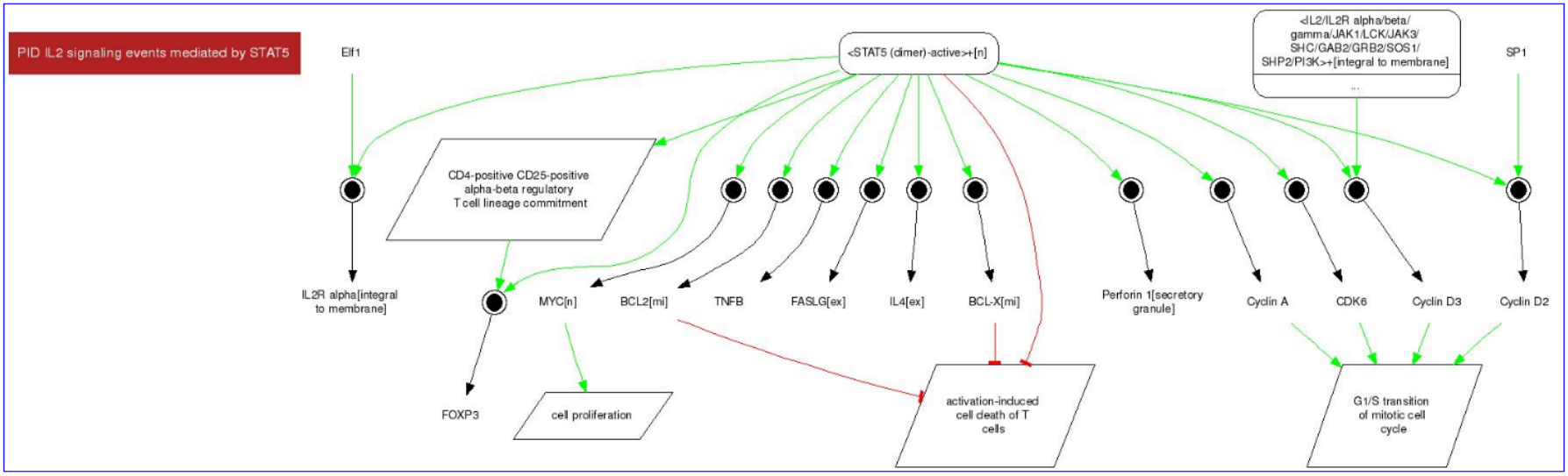
Molecule list | References

View graphic as: JPG | SVG | Silverlight-beta | Help

Save code as: XML | BioPAX | Help

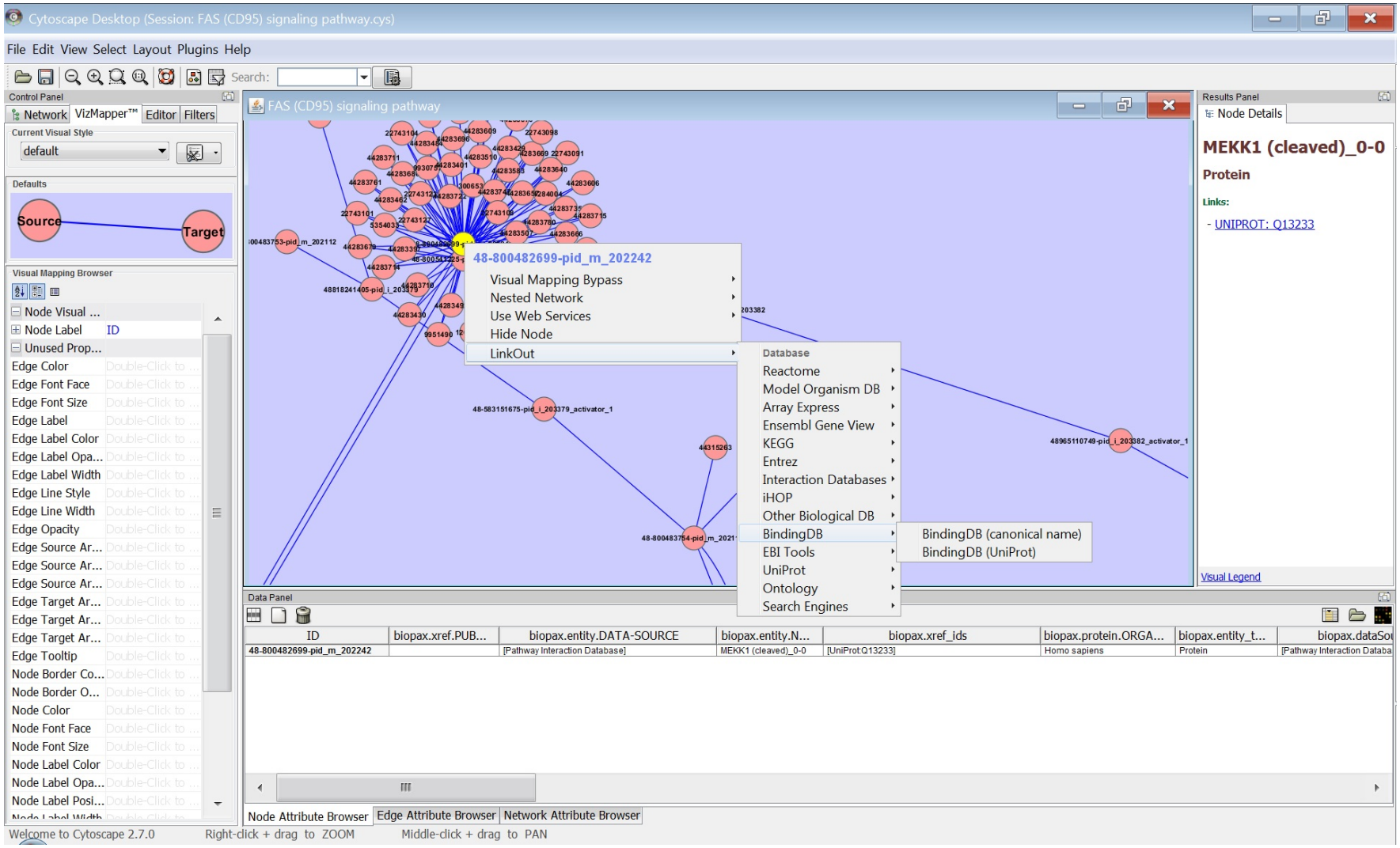
A key to the image icons appears below the pathway.

Click on a biomolecule or interaction for additional information.



Cytoscape Pathways and BindingDB

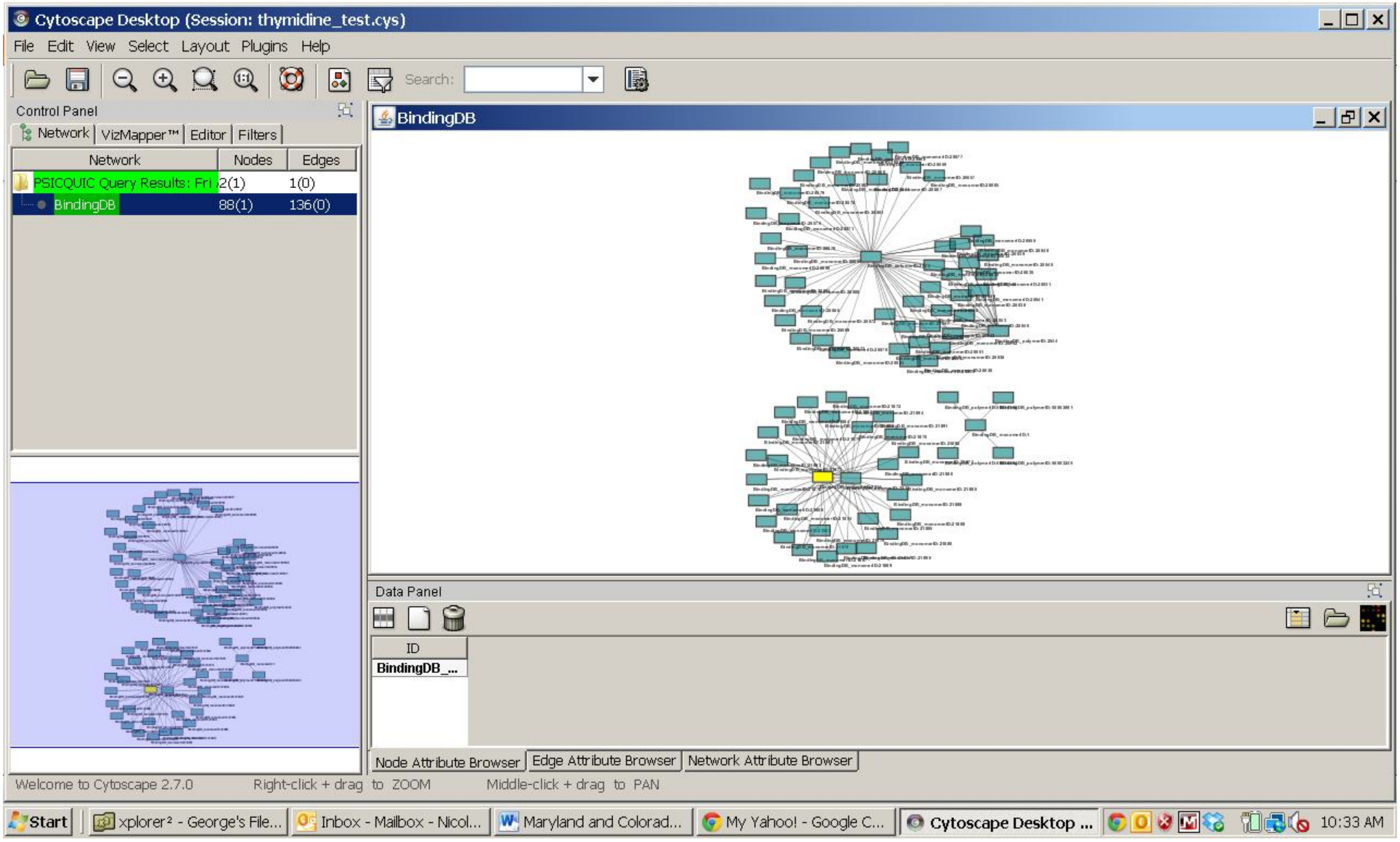
Proteins to Small Molecules and Affinities



With Drs. Trey Ideker and Michael Smoot, UCSD

Cytoscape to BindingDB via PSICQUIC Server

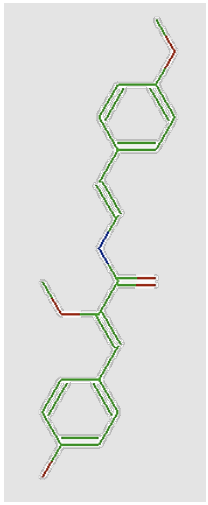
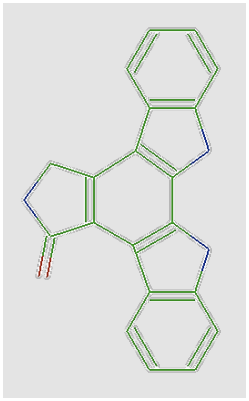
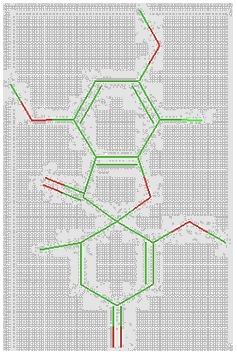
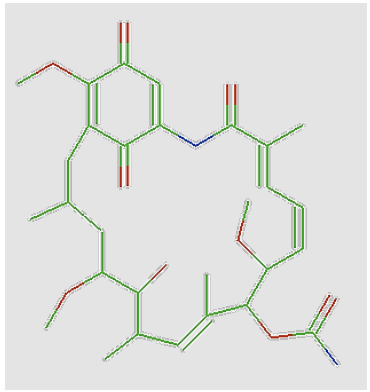
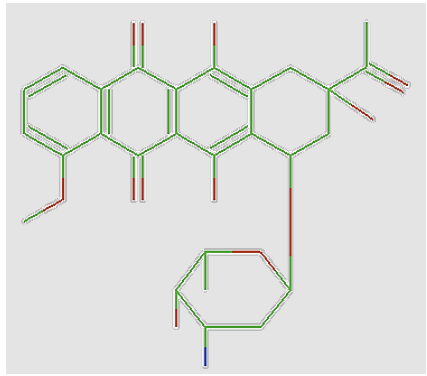
Preliminary Implementation



Marine Natural Products

MarinLit Compounds in BindingDB

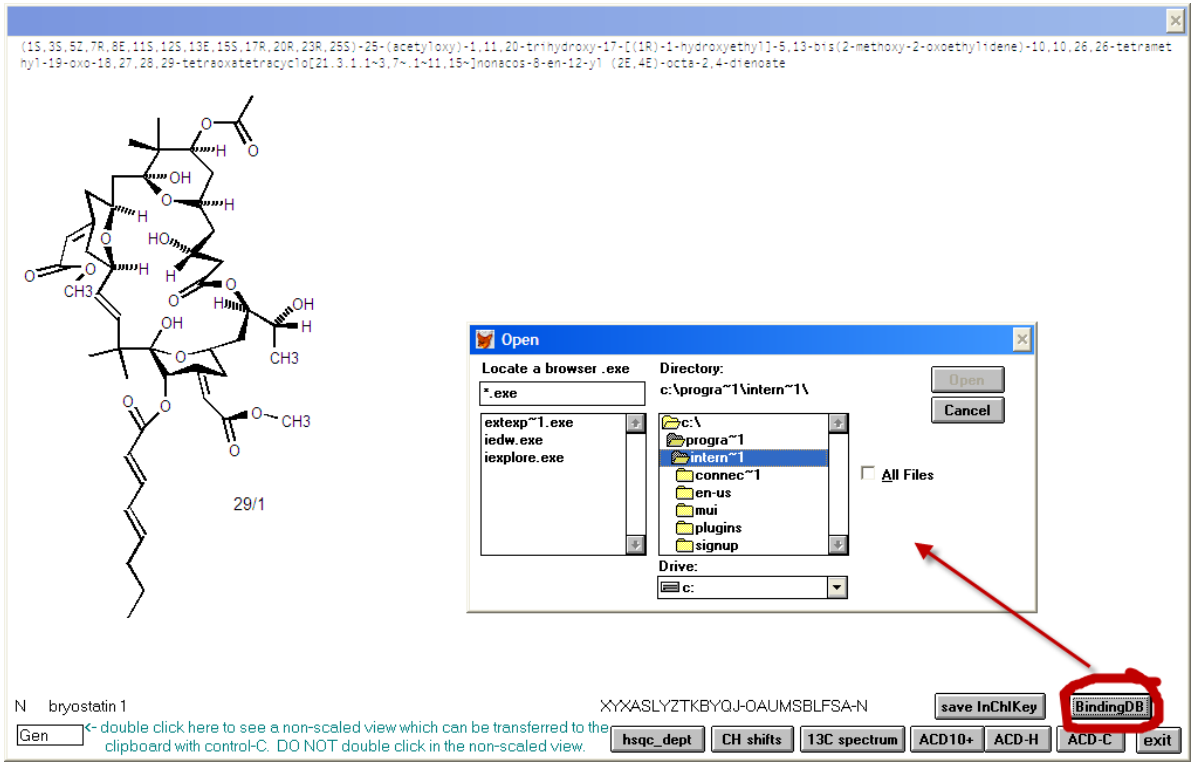
400 exact matches in BindingDB
5200 similarity matches >0.85

Hyperlinks from MarinLit to BindingDB

Initial Implementation

Dr. John Blunt, U. Canterbury, NZ



Prepared Validation Sets

For computational chemists

Criteria

- Congeneric compound series from one lab
- At least one representative complex in the PDB
- Range of affinities

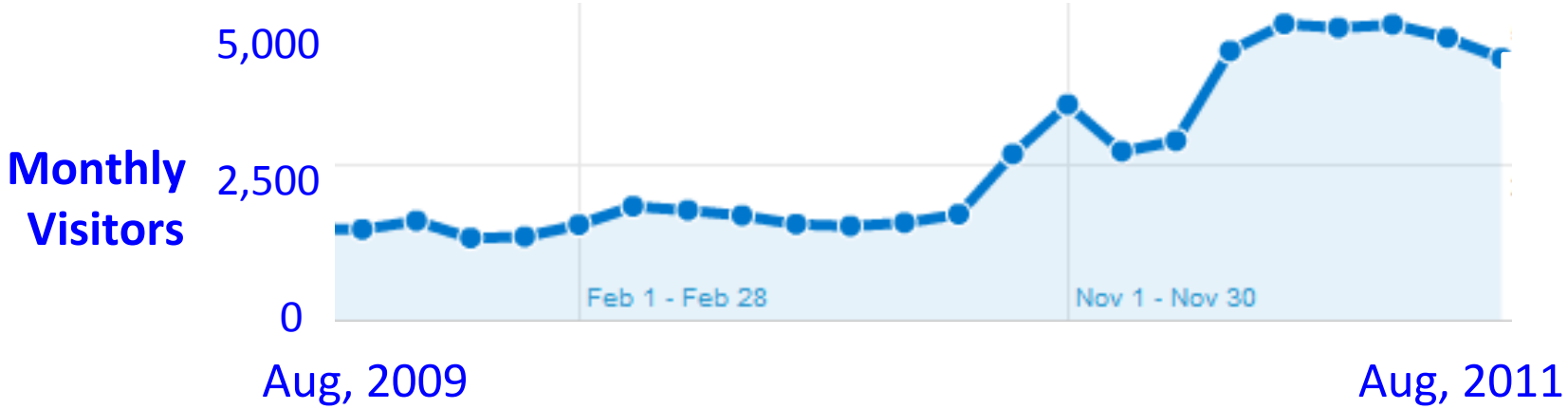
Features

- Easily downloaded (SDfile)
- Comments invited on each dataset

11-beta-Hydroxysteroid Dehydrogenase 1 (11-beta-HSD1) Add/Read (0)

	Compounds	Affinity (nM)	Structures (PDB ids)	Data Source(Articles)	SDfile	User Comments
Set 1	19	6.3 - 1000	2RBE	Bioorg Med Chem Lett. 2007 17:6056-61	Download	Add/Read (0)
Set 2	24	3 - 372	3BYZ	J Med Chem. 2008 51:2933-43	Download	Add/Read (0)
Set 3	30	3 - 6500	3BZU 3EY4	J Med Chem. 2008 51:7953-67 J Med Chem. 2007 50:429-32	Download	Add/Read (0)

Increased Usage and Stable Evaluations



User Survey

	Overall Impression	Scope of Data	Data Quality
2011	2.0	1.9	2.0
2008	1.7	2.0	1.9

(1=best, 5=worst)

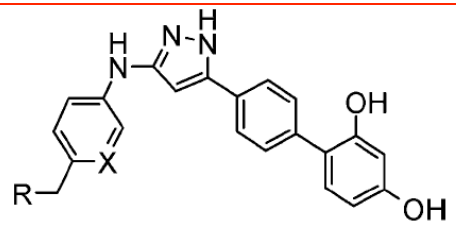
The Curation Problem

Labor, Errors and Cost

Chemist's computer
machine readable

Published Article
not machine readable

Database
machine readable

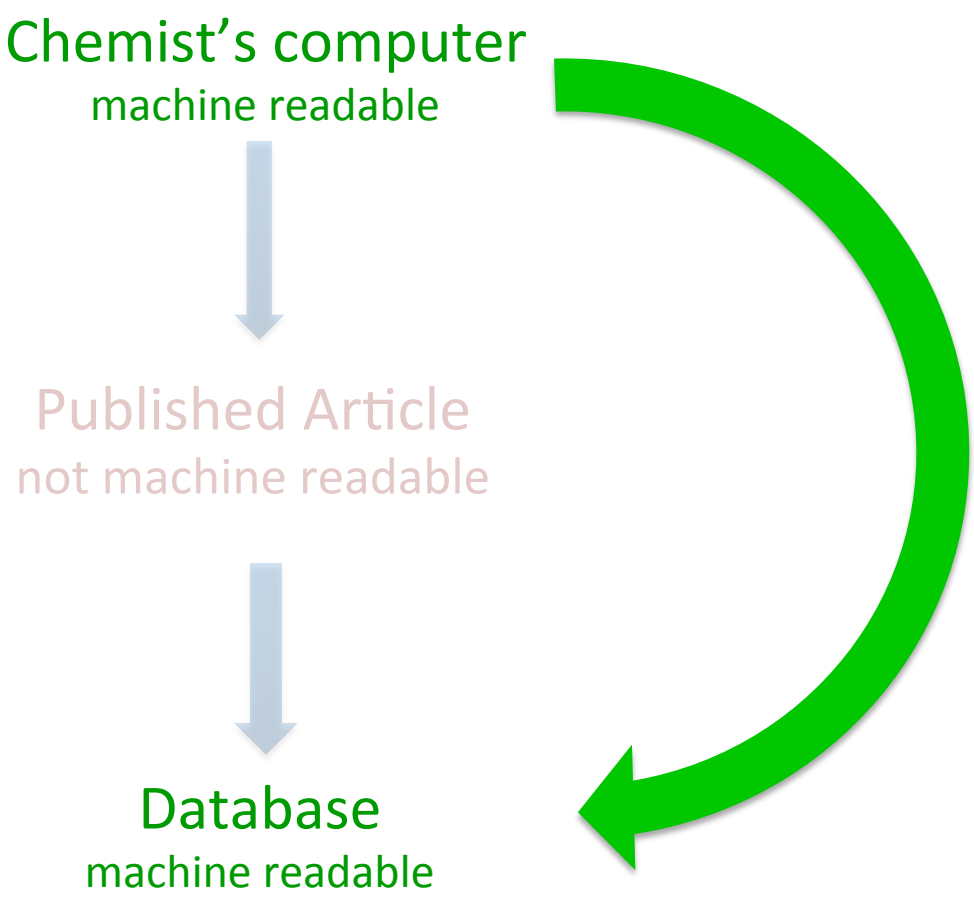


cmpd	X	R	K_i[a] (nM)	EC_{50}[b] (nM)
1			301	NA[c]
2			0.5	550
3			1.20	500
10a	CH	*iso*-propylamino	0.36	70
10b	CH	pyrrolidinyl	0.18	41
10c	CH	cyclopropylamino	0.38	60
10d	CH	dimethylamino	0.19	58
10e	N	piperidinyl	0.30	150
10f	N	*sec*-butylamino	0.20	230
10g	N	cyclopropylmethylamino	0.80	165

JMC 50:5253, 2007

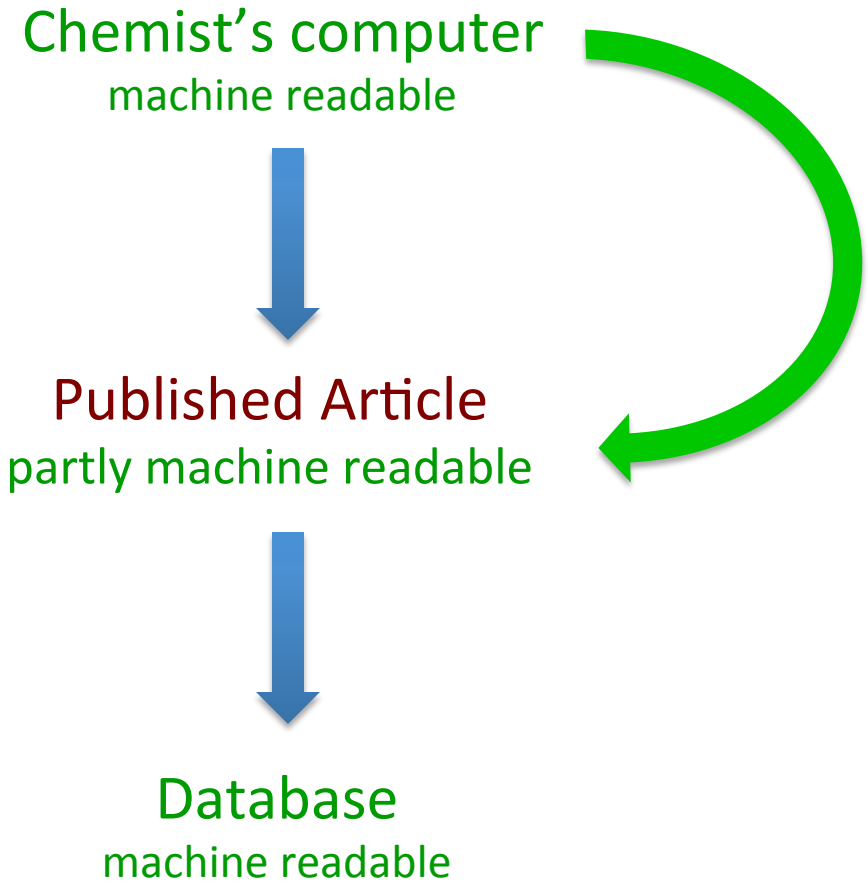
One Vision for Curation:

Eliminate It



A Second Vision for Curation

Keep Journals as Quality Filter



A Simple Step

Machine-Readable Molecules in Articles

Small molecules

Compound	InChIKey	PubChemID	InChi or SMILES string
1a			
1b			
2			
GK501			

A Simple Step

Machine-Readable Molecules in Articles

Proteins

Protein	GenBank ID	UniProtKB ID	FASTA Seq
hAR			
hER			

Compound	InChIKey	PubChemID	InChi or SMILES string
1a			
1b			
2			
GK501			

Molecule Interaction Matrix

A Bridge Too Far?

IC50	1a	1b	2	GK501
hAR				
hER				

Acknowledgements

NSF, NIH

BindingDB Team

Tiqing Liu

George Nicola

Linda Hwang

(Xi Chen)

(Yuhmei Lin)

(Ming Liu)

BindingDB Collaborators and Advisors

Phil Bourne (PDB,UCSD)
Peter Rose (PDB,UCSD)
John Westbrook (PDB,Rutgers)
Lei Xie (PDB,UCSD)
Other PDB staff
John Overington (ChEMBL)
John Irwin (UCSF/Zinc)
Steve Bryant (PubChem)
John Blunt (MarinLit, U. Canterbury, NZ)
Andy McCammon (UCSD)
Shankar Subramaniam (Pathways, UCSD)
Ashok Dinasarapu (Pathways, UCSD)
Trey Ideker (Cytoscape, UCSD)
Mike Smoot (Cytoscape, UCSD)
Sandra Orchard (PsiMEX,EBI)
Henning Hermjakob (PsiMEX,EBI)
Samuel Kerrian (EBI)
Ajay Jain (UCSF)

ChemAxon, OpenEye